DOA and Pitch Estimation of Audio Sources using IAA-based Filtering

Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

DOA AND PITCH ESTIMATION OF AUDIO SOURCES USING IAA-BASED FILTERING

Jesper Rindom Jensen and Mads Græsbøll Christensen

Audio Analysis Lab, AD:MT, Aalborg University, Denmark, email: {jrj,mgc}@create.aau.dk

ABSTRACT

For decades, it has been investigated how to separately solve the problems of both direction-of-arrival (DOA) and pitch estimation. Recently, it was found that estimating these parameters jointly from multichannel recordings of audio can be extremely beneficial. Many joint estimators are based on knowledge of the inverse sample covariance matrix. Typically, this covariance is estimated using the sample covariance matrix, but for this estimate to be full rank, many temporal samples are needed. In cases with non-stationary signals, this is a serious limitation. We therefore investigate how a recent joint DOA and pitch filtering-based estimator can be combined with the iterative adaptive approach to circumvent this limitation in joint DOA and pitch estimation of audio sources. Simulations show a clear improvement compared to when using the sample covariance matrix and the considered approach also outperforms other state-of-the-art methods. Finally, the applicability of the considered approach is verified on real speech.

Index Terms— Direction-of-arrival, fundamental frequency, linearly constrained minimum variance, iterative adaptive approach, high resolution.

1. INTRODUCTION

Microphone arrays are often used for recording audio sources such as speech and musical instruments. It is well known that many short segments of such recordings are approximately periodic, and that two particularly interesting parameters describing them are the fundamental frequency, herein denoted as the pitch, and the direction-of-arrival (DOA) of the audio source in relation to the microphone array. The interest emanates from the fact that the pitch, the DOA or both parameters are key components in a substantial body of signal processing methods for compression, separation, enhancement, beamforming, automatic music transcription, music classification, localization, etc. Many of these methods are or can be utilized in several applications, including hands-free communication, smart homes, hearing aids, surveillance systems, and teleconferencing. This emphasizes the importance of knowing the pitch and DOA.

This research was funded by the Danish Council for Independent Research, grant ID: DFF 1337-00084, and the Villum Foundation.

Traditionally, the estimation of the pitch and DOA has been treated as two independent estimation problems. DOA estimation [1], for example, has been considered in the context of geophysics, radio astronomy, biomedical engineering, radar, and microphone arrays. Likewise, pitch estimation has also been widely and independently studied, resulting in various classes of methods based on, e.g., autocorrelation functions, filtering, subspaces, and statistical models [2]. However, in the recent years, joint estimation of the pitch and DOA have been an increasingly considered research topic. Some of the advantages of conducting joint estimation were outlined in [3]: 1) if a source parameter is similar for both sources in a two-source scenario, the sources are not resolvable with separate estimation, and 2) joint estimation potentially gives a higher estimation accuracy. The research in joint DOA and pitch estimation, has resulted in different joint estimation approaches. Examples are maximum likelihood based [3], subspace-based [4, 5], correlation-based [6], and filtering-based [7–9] methods.

The filtering-based approach, being considered in this paper, has proven particularly useful in scenarios with multiple sources, since no explicit assumption about the noise is needed [2]. However, in these filtering methods, the inverse covariance matrix of the observed signal is needed, which is problematic in practice. The reason is that the so-called spatio-temporal sample covariance matrix estimate is typically utilized, but for this to be full rank in the multichannel case, a large number of temporal samples is usually needed [3]. In other words, most previously proposed filtering methods for joint DOA and pitch estimation of audio sources have not been applicable on short multichannel observations. We, therefore, consider the application of an optimal filtering approach [9] for joint DOA and pitch estimation to audio sources. This approach combines LCMV filtering [8] with the iterative adaptive approach (IAA) [10]. More specifically, we extend on the work in [9] by 1) investigating the advantages of using the IAA covariance matrix estimate rather than the sample covariance matrix estimate both theoretically and through simulations, 2) comparing the IAA-based LCMV filtering method for joint DOA and pitch estimation with state-of-the-art methods, and 3) showing how the IAA-based LCMV filtering method can be applied on real audio recordings. While computational complexity of the considered method is relatively high, it can be lowered dramatically by

generalizing recently proposed implementations, e.g., [11] to the multichannel scenario.

The remainder of the paper is organized as follows: in Section 2, we introduce our signal model, and formulate the research problem. Then, in Section 3, we present the LCMV filter for joint DOA and pitch estimation, and we consider covariance estimation in Section 4. The experimental results are found in Section 5, and, finally, a discussion is found in Section 6.

2. SIGNAL MODEL AND PROBLEM FORMULATION

We consider the scenario, where K microphones are used for recording a desired signal added to a mixture of background noise and interfering sources. At time instance n for the k'th microphone, the recorded signal can be modeled as $y_k(n) = x_k(n) + v_k(n)$, where $x_k(n)$ is the desired signal, and $v_k(n)$ is a sum of the background noise and interfering sources. The topic of this paper is joint estimation of the pitch and DOA of periodic sources recorded using a microphone array, so, naturally, we assume the desired signal to be periodic. This has proven to be a good assumption for short segments of voiced speech and musical instrument signals [2]. The background noise can, for example, be sensor noise, while examples of interfering sources may be other directive periodic sources (fan, speech, etc.). By exploiting the periodicity assumption, and by assuming that the microphones are situated in relative close vicinity of each other, the observed signal can be further modeled as [3]

$$y_k(n) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_0(n-f_s\tau_k)} + v_k(n), \tag{1}$$

with L denoting the number of harmonics, $\alpha_l = A_l e^{j\phi_l}$ is the complex amplitude of the l'th harmonic, $A_l > 0$ and ϕ_l are the real amplitude and phase of the l'th harmonic, ω_0 is the pitch, f_s is the sampling frequency, and τ_k is the delay of the desired signal from microphone 0 to microphone k. Reverberation is not explicitly accounted for in the model, but it is partly represented by $v_k(n)$. The delay can be further modeled if we know how the microphones are situated in relation to each other. While any such microphone array structure can be considered, we assume a uniform linear array (ULA) structure herein. In this case, the delay is given by $\tau_k = k\frac{d\sin\theta}{c}$, where d is the spacing between two neighboring microphones, θ is the direction-of-arrival (DOA) of the desired source in relation to the array, and c is the propagation speed of the sound wave. With this model of the delay, the observed signal model can be rewritten as

$$y_k(n) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_0 n} e^{-jl\omega_s k} + v_k(n), \;\; \omega_s = \omega_0 f_s \tau_1. \tag{2}$$

In practice, we wish to estimate the DOA and pitch from N consecutive samples from K microphones. Stacking these observations in vectors helps us in the derivation of optimal filtering methods for solving the estimation problem. The observed signals can, e.g., be organized as $[\mathbf{Y}(n)]_{km} = y_k(n-m)$, for $k = 0,\ldots,K-1$ and $m = 0,\ldots,N-1$, and $[\cdot]_{km}$ denotes the (k,m)'th entry of a matrix. In the considered estimation method, subblocks of the observed signal are also used, and these are defined as

$$\mathbf{Y}_k(n) = \sum_{l=1}^{L} \alpha_l(n)\mathbf{z}_s(l\omega_s)\mathbf{z}_t^T(l\omega_0) + \mathbf{V}_k(n), \tag{3}$$

where $[\mathbf{Y}_k(n)]_{pm} = y_{k+p}(n-m)$, for $p = 0,\ldots,P-1$ and $m = 0,\ldots,M-1$, $[\mathbf{V}_k(n)]_{pm}$ is defined similarly to $[\mathbf{Y}_k(n)]_{pm}$, $\alpha_l(n) = e^{jl\omega_0 n}$, $[\mathbf{z}_s(l\omega_s)]_p = e^{-jlp\omega_s}$, for $p = 0,\ldots,P-1$, where $[\cdot]_p$ denotes the p'th entry of a column vector, $[\mathbf{z}_t(l\omega_0)]_m = e^{-jlm\omega_0}$, for $m = 0,\ldots,M-1$, P is the spatial subblock length, M is the temporal subblock length, and $(\cdot)^T$ denotes the matrix transpose. Stacking the columns of the matrices, yields

$$\mathbf{y}_k(n) = \text{vec}\{\mathbf{Y}_k(n)\} = \sum_{l=1}^{L} \alpha_l(n)\mathbf{z}_l + \mathbf{v}_k(n), \tag{4}$$

with $\text{vec}\{\cdot\}$ denoting the operator that stacks the columns of a matrix, $\mathbf{v}_k(n) = \text{vec}\{\mathbf{V}_k(n)\}$, and $\mathbf{z}_l = \mathbf{z}_s(l\omega_s) \otimes \mathbf{z}_t(l\omega_0)$, where $\otimes$ is the Kronecker product operator.

3. OPTIMAL FILTERING METHOD

As initially shown in [3], optimal filtering can be used to solve the joint DOA and pitch estimation problem. We briefly present the highlights of this approach in the following section. First of all, we define the spatio-temporal filtering operation as $z_k(n) = \mathbf{h}^H\mathbf{y}_k(n)$, where $\mathbf{h}$ is the spatio-temporal filtering vector, and $(\cdot)^H$ denotes the complex conjugate transpose of matrix or vector. The idea is then to design a filter that passes the desired signal, in this case a harmonic signal, undistorted, while reducing the noise as much as possible. Mathematically, this design problem is equivalent to

$$\min_{\mathbf{h}} \mathbf{h}^H\mathbf{R}_\mathbf{y}\mathbf{h} \;\; \text{s.t.} \;\; \mathbf{h}^H\mathbf{z}_l = 1, \;\; \text{for} \;\; l = 1,\ldots,L,$$

where $\mathbf{R}_\mathbf{y} = \text{E}[\mathbf{y}_k(n)\mathbf{y}_k^H(n)]$, and $\text{E}[\cdot]$ denotes the mathematical expectation. The well-known solution to this quadratic optimization problem can be obtained using Lagrange multipliers and is given by

$$\mathbf{h} = {\mathbf{R}_\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H{\mathbf{R}_\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1}, \tag{5}$$

with $\mathbf{1} \in \mathbb{R}^L$ denoting a vector of ones, and $\mathbf{Z} = [\mathbf{z}_1 \;\; \cdots \;\; \mathbf{z}_L]$. The DOA and pitch are then estimated jointly, by designing the optimal filter for different candidate DOAs and pitch frequencies, and maximizing the output power of the filter over

these candidates. This can also be written as

$$\{\widehat{\omega}_0, \widehat{\theta}\} = \arg \max_{\{\omega_0,\theta\}\in\Omega\times\Theta} \mathbf{h}^H \mathbf{R}_{\mathbf{y}} \mathbf{h}, \tag{6}$$

where Ω and Θ denote the sets of candidate DOAs and pitch frequencies, respectively. In practice, the covariance matrix of the observed signal $\mathbf{R}_{\mathbf{y}}$ is unknown and has to be estimated. Care should be taken, though, that the estimate is invertible due to the expression in (5).

4. COVARIANCE ESTIMATION

The traditional way of estimating the covariance matrix is to use a spatio-temporal sample covariance matrix estimate. This outer product estimate of $\mathbf{R}_{\mathbf{y}}$ is given by [3]

$$\widehat{\mathbf{R}}_{\mathbf{y}} = \sum_{k=0}^{K-P} \sum_{m=0}^{N-M} \frac{\mathbf{y}_k(n-m)\mathbf{y}_k^H(n-m)}{(K-P+1)(N-M+1)}. \tag{7}$$

For this estimate of the covariance matrix to be invertible, we must require that $(K-P+1)(N-M+1) \geq MP$. Typically, the number of microphones is small, and $K \ll N$, so to achieve a reasonable spatial resolution with the resulting estimator, we choose $P = K$. In this case, the inequality can be rewritten as $M \leq \frac{N+1}{K+1}$. That is, M should be very small or a large amount of temporal samples N is needed if K is relatively large.

To avoid these limitations, we consider the use of the IAA [10] in conjunction with the optimal filtering method for joint DOA and pitch estimation. In [10], the IAA was applied for spectral amplitude estimation in two dimensions, namely range and DOA. In the following, we show how this principle can be used for spatio-temporal spectral amplitude and covariance estimation. First, we denote an amplitude of a spatio-temporal frequency component of interest by $\alpha_{\gamma',\psi'}$, where γ' is a frequency index, and ψ' is a spatial frequency index corresponding to the DOA. By utilizing the covariance matrix model [12], the noise covariance matrix can then be approximated as

$$\mathbf{Q}_{\gamma',\psi'} \approx \widetilde{\mathbf{R}}_{\mathbf{y}} - |\alpha_{\gamma',\psi'}|^2 \mathbf{z}_{\gamma',\psi'} \mathbf{z}_{\gamma',\psi'}^H, \tag{8}$$

$$\widetilde{\mathbf{R}}_{\mathbf{y}} = \sum_{\gamma=1}^{\Gamma} \sum_{\psi=1}^{\Psi} |\alpha_{\gamma,\psi}|^2 \mathbf{z}_{\gamma,\psi} \mathbf{z}_{\gamma,\psi}^H, \tag{9}$$

γ and ψ denote frequency and spatial frequency indices, respectively, Γ is the number of frequency grid points utilized in the IAA, and Ψ is the number of spatial frequency grid points utilized in the IAA. The matrix $\widetilde{\mathbf{R}}_{\mathbf{y}}$ can be seen as an estimate of the signal covariance matrix, and $\mathbf{z}_{\gamma,\psi} = \mathbf{z}_{\mathrm{s}}(\psi) \otimes \mathbf{z}_{\mathrm{t}}(\gamma)$, $[\mathbf{z}_{\mathrm{t}}(\gamma)]_n = e^{-jn\omega_\gamma}$, for $n = 0, \ldots, N-1$, $[\mathbf{z}_{\mathrm{s}}(\psi)]_k = e^{-jk\omega_{\mathrm{s},\psi}}$, for $k = 0, \ldots, K-1$, $\omega_\gamma = \frac{\gamma-1}{\Gamma} 2\pi$ denotes the frequency corresponding to the γ'th grid point, and $\omega_{\mathrm{s},\psi} = \frac{\psi-1}{\Psi} 2\pi$ denotes the spatial frequency corresponding to the grid point ψ.

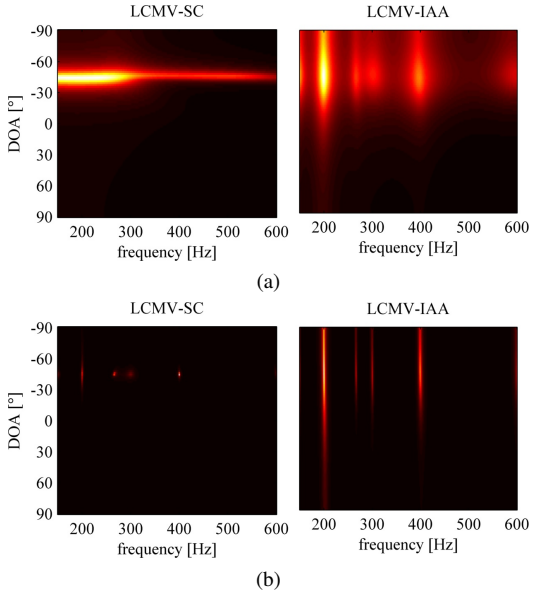

Fig. 1. Plots of the cost-functions for the LCMV-IAA and LCMV-SC methods when applied on a synthetic, multichannel, periodic signal for (a) $N = 20$ and (b) $N = 80$.

The IAA is then used to estimate the amplitude $\alpha_{\gamma',\psi'}$ through minimization of a weighted least squares (WLS) cost-function defined as

$$J_{\mathrm{WLS}} = \left[\mathbf{y}(n) - \alpha_{\gamma',\psi'} \mathbf{z}_{\gamma',\psi'}\right]^H \mathbf{Q}_{\gamma',\psi'} \left[\mathbf{y}(n) - \alpha_{\gamma',\psi'} \mathbf{z}_{\gamma',\psi'}\right],$$

where $\mathbf{y}(n) = \mathrm{vec}\{\mathbf{Y}(n)\}$. If we minimize J_{WLS} with respect to the unknown amplitude $\alpha_{\gamma',\psi'}$, we get the following closed-form estimate [10]

$$\hat{\alpha}_{\gamma',\psi'} = \left(\mathbf{z}_{\gamma',\psi'}^H \widetilde{\mathbf{R}}_{\mathbf{y}}^{-1} \mathbf{z}_{\gamma',\psi'}\right)^{-1} \mathbf{z}_{\gamma',\psi'}^H \widetilde{\mathbf{R}}_{\mathbf{y}}^{-1} \mathbf{y}(n). \tag{10}$$

We note that the amplitude estimate depends on the estimate of the covariance matrix and vice versa, so these are estimated by iterating between (9) and (10), hence the method is termed the IAA. While the IAA has historically been used for amplitude spectrum estimation, we here utilize it for estimation of the covariance matrix of the observed signal herein. As opposed to the sample covariance matrix estimate, this estimate is formed from a single observation, $\mathbf{y}(n)$, while also being full-rank. This enables us to choose $M = N$ and $P = K$, but of course it is computationally more complex to obtain this estimate than the sample covariance matrix estimate. The IAA is initialized with $\widetilde{\mathbf{R}}_{\mathbf{y}} = \mathbf{I}$, and, typically, 10-15 iterations is sufficient to achieve convergence.

5. EXPERIMENTAL RESULTS

First, we compare the considered method (LCMV-IAA), with LCMV filtering based on the spatio-temporal sample covariance matrix estimate (LCMV-SC). For this experiment, we used a synthetic, periodic signal with $L = 4$, $\theta = -45°$,

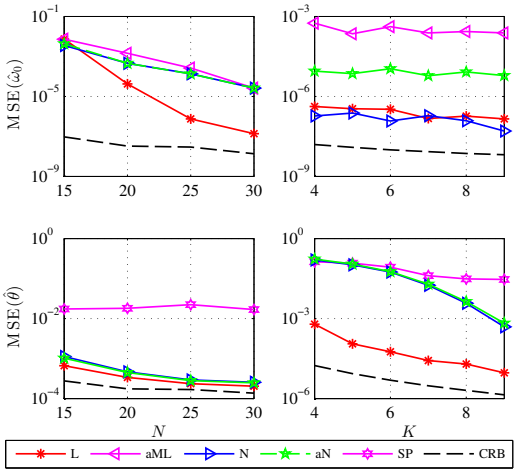

Fig. 2. MSEs of pitch and DOA estimates for different N's and K's for the considered IAA-based LCMV and state-of-the-art methods for joint DOA and pitch estimation.

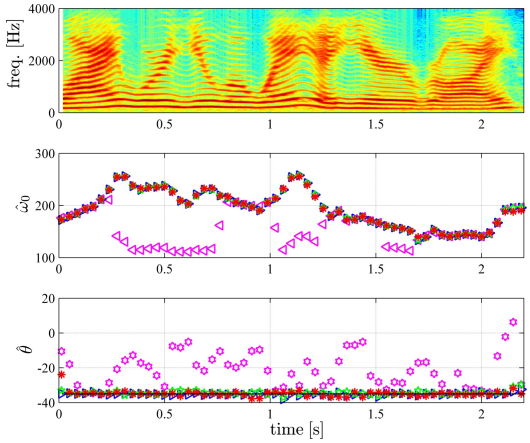

Fig. 3. Plots of (top) the spectrogram of a speech signal, and estimates of its (middle) pitch and (bottom) DOA. The legend for the plot is found in Fig. 2

$f_0 = 200$ Hz, $f_s = 4$ kHz, the harmonics each had unit amplitudes, and white noise was added to each channel at an SNR of 30 dB. The remaining simulation set up was: $K = 3$, $P = 3$, $N = 20$, $M = \lfloor \frac{N+1}{K+1} \rfloor$, $\Gamma = 512$, $\Psi = 128$, 10 IAA iterations were used, $c = 343$ m/s, and $d = 0.04$ m. With this setup, we implemented the LCMV-IAA and LCMV-SC methods, and evaluated their respective cost-functions for different candidate DOAs and pitch frequencies, and the results are shown in Fig. 1a. Clearly, the LCMV-SC method is not able to estimate the DOA and pitch accurately from this small segment, due to a poor frequency resolution, as opposed to the LCMV-IAA method. If we then raise the number of temporal samples to $N = 80$, we get the results in Fig. 1b. With this increased number of samples, the LCMV-SC method is now able to obtain a reasonable DOA and pitch estimate, and it even seems that it has higher spatial resolution than LCMV-IAA. This indicates that the LCMV-IAA method should be applied on short segments, while LCMV-SC should be applied on longer segments.

The estimation accuracy of the considered method was also evaluated in terms of mean squared errors (MSEs) in series of Monte-Carlo simulations. In these simulations, we compared the considered method (L), with the multichannel pitch estimator in [13] (aML), the exact and asymptotic joint NLS estimators in [3] (N and aN), and the SRP-PHAT method for DOA estimation [14] (SP). Note that the considered, LCMV method was implemented with the IAA covariance matrix estimate, since relatively small N's are considered. The methods were compared on different scenarios with a single harmonic signal as the desired signal added with a mixture of an interfering source and white Gaussian noise. In these experiments, a harmonic signal with 4 unit amplitude harmonics were used with a the pitch being sampled from $\mathcal{U}(250 \text{ Hz}, 300 \text{ Hz})$ and the DOA being sampled from $\mathcal{U}(15°, 35°)$ in each Monte-Carlo simulation. The interfering sinusoid also had unit amplitude, and the white noise was added at a 30 dB SNR in relation to the desired signal. Besides that, $f_s = 4$ kHz, $c = 343$ m/s, and $d = 0.04$ m, $\Gamma = 256$, and $\Psi = 64$. With this setup, we first considered a scenario, where the interfering source had the same DOA as the desired signal, but with a frequency equal to $f_i = f_0 + 40$ Hz. In this scenario, the pitch and DOA was estimated with different N's and $K = 2$, and for each N, the MSE of the estimates was found from 100 Monte-Carlo simulations. In another scenario, the interfering sinusoid had the same frequency as the pitch of the desired signal, but with a DOA of $\theta_i = 80°$. For this scenario, $\Gamma = 256$ and $\Psi = 128$, and the MSEs were measured for different K's with 100 Monte-Carlo simulations for each K, and N was fixed to 20. The results obtained from these scenarios are shown in Fig. 2. We see that for $N \geq 20$ the L method outperforms all the other methods for pitch estimation, and, similarly, all other methods are outperformed for all considered K's in terms of DOA estimation. Otherwise, the performance of the L method is comparable to that of the NLS method.

In the final experiment, the considered method was evaluated on a real-signal. The signal used was a 2.4 seconds long, single-channel speech signal, which was resynthesized spatially, using an online available room impulse response generator [15]. The RIR generator was set up as follows: $c = 343$ m/s, $f_s = 8$ kHz, the microphones of a ULA was located at $[2+d(k-\frac{K-1}{2})]$ m$\times$0.1 m$\times$1.5 m for $k = 1, 2, 3$, $d = 0.04$ m, the source was located at $\theta = -35°$ and $r_c = 2$ m, the room dimensions was 4 m$\times$4 m$\times$3 m, the length of the RIRs was 2,048, the microphone types was omnidirectional, and the reflection order was 0. With this setup, we generated the spatial-temporal data on which the aforementioned methods were applied on consecutive frames of length $N = 50$ of the multichannel signal. The estimators were implemented by assuming that $L = 6$ and with $\Gamma = 128$ and $\Psi = 64$. In the SRP-PHAT method, we used an FFT length of 256 and inte-

grated over frequencies in the interval $[150 \text{ Hz}, f_s/2]$. Moreover, due to pitch halfing/doubling (i.e., pitch estimates being approximately half/twice the true pitch) as an effect of choosing a fixed model order, the pitch estimates were smoothed using the method in [16] with the bonus parameter set to 150. The results in Fig. 3 were obtained with this setup. Comparing the obtained pitch estimates with the spectrogram of the speech signal, it is difficult to judge which method has the highest accuracy, but the L, N, and aN methods clearly outperform the aML method. The L, N, and aN estimators seem to obtain pitch estimates in the correct frequency region. Regarding DOA estimation, the considered method shows similar accuracy compared to the N and aN methods, but it clearly seems to outperform the other SP method.

6. DISCUSSION

We considered joint estimation of the DOA and pitch of a periodic source captured using a microphone array in this paper. The joint estimation of these parameters have only been considered for less than a decade, and only relatively few of such methods exist, with some examples being [3–9]. Some of the recent approaches are based on the inverse spatio-temporal covariance matrix, and this covariance is most often replaced by the sample covariance matrix estimate in practice. However, for this estimate to be full rank, we need a large number of temporal samples, and this becomes even more pronounced by raising the numbers of sensors. If the signal of interest is highly non-stationary, this will clearly be a huge limitation of these methods. Herein, we consider the idea of combining a recent, filtering-based joint DOA and pitch estimator [3] with the IAA [10] for covariance matrix estimation, an idea that was spawned in [9]. We extend the work in [9] by investigating the advantages of using the IAA covariance matrix estimate rather than the sample covariance matrix estimate both theoretically and through simulations. We also compare the IAA-based LCMV filtering method for joint DOA and pitch estimation with state-of-the-art methods, and show its applicability to multichannel audio recordings.

REFERENCES

[1] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.

[2] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[3] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 21, no. 5, pp. 923–933, May 2013.

[4] L. Y. Ngan, Y. Wu, H. C. So, P. C. Ching, and S. W. Lee, "Joint time delay and pitch estimation for speaker localization," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2003, vol. 3, pp. 722–725.

[5] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and multi-pitch estimation based on subspace techniques," *EURASIP J. on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–11, Jan. 2012.

[6] M. Képesi, L. Ottowitz, and T. Habib, "Joint position-pitch estimation for multiple speaker scenarios," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, May 2008, pp. 85–88.

[7] J. Dmochowski, J. Benesty, and S. Affes, "Linearly constrained minimum variance source localization and spectral estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 8, pp. 1490–1502, Nov. 2008.

[8] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Joint DOA and fundamental frequency estimation methods based on 2-d filtering," in *Proc. European Signal Processing Conf.*, Aug. 2010, pp. 2091–2095.

[9] Z. Zhou, M. G. Christensen, J. R. Jensen, and H. C. So, "Joint DOA and fundamental frequency estimation based on relaxed iterative adaptive approach and optimal filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013.

[10] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. A. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 5–20, Feb. 2010.

[11] G.-O. Glentis and A. Jakobsson, "Superfast approximative implementation of the IAA spectral estimate," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 472–478, Jan. 2012.

[12] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Education, Inc., 2005.

[13] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 409–412.

[14] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays - Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., chapter 8, pp. 157–180. Springer-Verlag, 2001.

[15] E. A. P. Habets, "Room impulse response generator," Tech. Rep., Technische Universiteit Eindhoven, 2010, Ver. 2.0.20100920.

[16] H. Ney, "A dynamic programming algorithm for non-linear smoothing," *Signal Process.*, vol. 5, no. 2, pp. 163–173, Mar. 1983.